UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS

TESIS

“Aplicación de redes neuronales convolucionales para el etiquetado de imágenes automático para personas con impedimentos visuales”

PLAN DE TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERIO DE SISTEMAS

ELABORADO POR

SALAZAR VALVERDE, FREDDY DICK

ASESOR

ZELA MORAYA, WESTER

2019

DEDICATORIA

Esta tesis se la dedico a mis seres queridos, por todo el apoyo y confianza que depositan en mí, para lograr mis metas trazadas.

AGRADECIMIENTO

Agradezco a mis padres por todo el amor y cariño que tienen hacia a mí, por las enseñanzas de la vida, he aprendido mucho de ellos, a mis hermanos por ser como son y todos mis seres queridos que son la motivación para realizar esta tesis.

RESUMEN

El presente trabajo de investigación pretende mejorar la calidad de vida de las personas con alguna discapacidad visual, mediante un método de clasificación de imágenes usado para reconocer objetos, basado en aprendizaje profundo y el uso de redes neuronales convolucionales, que, a partir de una gran cantidad de imágenes para entrenamiento, para luego obtener un modelo que permite predecir o etiquetar una imagen, a su vez se ´tiene una versión móvil lo que hace de muy fácil acceso y bajo costo.

ABSTRACT

This research work aims to improve the quality of life of people with some visual impairment, through an image classification method used to recognize objects, based on deep learning and the use of convolutional neural networks, which from a large amount of images for training, to then obtain a model that allows to predict or label an image, in turn it has a mobile version which makes it very easy to access and low cost.

Contenido

CAPÍTULO I INTRODUCCIÓN

1.1 GENERALIDADES

En el mundo los avances tecnológicos que se están viviendo en los últimos años son impresionantes, la transformación digital, la aparición de nuevos dispositivos electrónicos como el teléfono inteligente, la computadora, laptop y otros dispositivos electrónicos, los cuales traen grandes beneficios funcionales para la humanidad son herramientas de vital importancia en las labores diarias, sin embargo, el uso excesivo de estas herramientas puede traer consecuencias a la salud, uno de estos que es de gran importancia es la dificultad visual, que, en estos últimos años se ha incrementado en gran medida, lo que hace más probable que sufran alguna discapacidad visual en el futuro, por ende aumenta el riesgo de la salud visual, a su vez también han surgido avances en otros campos como la inteligencia artificial, el internet de las cosas, o el internet del todo, que están siendo aplicados para diversos ámbitos en la salud como para detectar enfermedades en fase temprana a través de la inteligencia artificial, identificando patrones encontrados en los pacientes que sufren alguna enfermedad, así como herramientas para controlar el estado de la salud del cuerpo en el instante, latidos del corazón, entre otras métricas, que son de vital importancia para Personas que sufren alguna enfermedad o discapacidad, así también la creciente oferta de aplicaciones al alcance de la mano como las diseñadas para los dispositivos

móviles, también existen algunos avances en los dispositivos clásicos de salud como las gafas, los lentes de contacto en el campo de la salud visual.

Ante este creciente avance tecnológico se han reportado según la Organización Mundial de la Salud el aumento de las cifras de Personas con discapacidad visual, ya que el uso excesivo de los teléfonos inteligentes, laptops, computadores, al cual estar expuestos por demasiado tiempo son dañinos a largo plazo, el Perú no es ajeno a ello gran parte de la población hace uso de estos dispositivos sin medir las consecuencias, ya que estas se ven reflejadas a largo plazo; y al parecer este crecimiento de Personas con alguna discapacidad visual no disminuirá en un futuro cercano, sin embargo los países están implementando políticas para hacerle frente, haciendo estudios masivos de prevención e implementado reformas políticas, ya que en su mayoría la discapacidad visual se puede prevenir según un informe de la Organización Mundial de la Salud.

A su vez existen alternativas de solución como son el uso de gafas, lentes de contacto o cirugías, que son soluciones parciales y la cirugía la cual tiene un costo elevado y es de difícil acceso.

No muchos saben qué es lo que sucede o pasa en el interior de un hogar de una Persona con discapacidad visual, cómo es que viven su día a día y qué es lo que en verdad les afecta y no les permite tener una mejor calidad de vida.

Ante esto se plantean varias interrogantes, cómo una Persona con discapacidad visual va hacia el baño y se cepilla, o se toma un baño, una labor

o actividad que al parecer es bastante sencilla, en verdad para ellos no lo es, o por si se le cae un objeto, deben buscarlo con el tacto.

Por lo antes expuesto, esta tesis propone, una herramienta para Personas con alguna discapacidad visual, haciendo uso de inteligencia artificial para poder reconocer objetos con el apoyo de la visión artificial, que no pueden reconocer cosas u objetos, o aquellas que no pueden ver, además es una herramienta de bajo costo y puede estar al alcance de cualquier Persona en un teléfono inteligente.

Esta herramienta permitirá Personas con alguna discapacidad visual a poder reconocer objetos, que quizás por la edad, ya que la mayoría de Personas con discapacidad visual son mayores de edad y también sufren o no pueden recordar con facilidad, no puedan recordar donde dejaron algún objeto importante, que requiere, para esto ellos deben buscar con el tacto el objeto que necesitan.

Para poder etiquetar estos objetos o imágenes de objetos y poder ayudar a las Personas con discapacidad visual usamos inteligencia artificial.

En general, la tarea de etiquetado efectivo de imágenes consiste en dos etapas, que involucran el etiquetado de imagen inicial y el posterior refinamiento de la etiqueta, estas dos etapas son claves para ayudar a los usuarios a acceder a las imágenes, pero no es fácil realizarla con precisión.

Para ayudar en esta tarea recurrimos a las técnicas de aprendizaje profundo, el cual es un campo perteneciente a la inteligencia artificial cuyo objetivo es el estudio y construcción de sistemas de cómputo capaces de aprender a partir

de la experiencia, inspirándose ligeramente en algunos principios del funcionamiento del cerebro animal. Existen dos paradigmas principales en la investigación de aprendizaje profundo, es decir, aprendizaje supervisado y aprendizaje no supervisado. El primer paradigma es una técnica para deducir una función a partir de datos de entrenamiento cuyo fin es crear una función capaz de predecir el valor correspondiente a cualquier objeto de entrada válida después de haber visto una serie de ejemplos, los datos de entrenamiento. Para ello, tiene que generalizar a partir de los datos presentados a las situaciones no vistas previamente. En cambio, el segundo paradigma los algoritmos pueden aprender automáticamente las representaciones conceptuales, como las caras de gatos y cuerpos humanos a partir de datos no marcados con metodologías no supervisadas.

En este documento analizaremos el etiquetado de imágenes a través del enfoque del aprendizaje supervisado haciendo uso de las redes neuronales convolucionales, que es un tipo de red neuronal artificial donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas en la corteza visual de un cerebro biológico. Esta red neuronal tiene un funcionamiento similar al proceso de sinapsis del cerebro ya que posee varias capas que a su vez tienen nodos que se conectan con los nodos de las siguientes capas. Esta estructura permitirá´ entrenar un conjunto de imágenes con sus respectivas etiquetas y mediante el uso de métricas se evaluará la eficiencia de este modelo es decir que tanto ha predicho los resultados de las etiquetas de las imágenes. El resto del trabajo esta´ organizado de la siguiente manera. El capítulo 2 se hace una breve revisión bibliográfica, se hace una

exposición de los conceptos necesarios para entender los modelos de aprendizaje supervisado y el etiquetado de imágenes. El capítulo 3 se presenta la fuente de datos y se describe las imágenes y las etiquetas usados en esta investigación. El capítulo 4 se presenta el modelo de red neuronal convolucional además se describirá las entradas, procesos y salidas. En el capítulo 3 modelo solución se presenta las conclusiones de la investigación del uso de redes neuronales convolucionales en el etiquetado de imágenes.

1.2 DESCRIPCIÓN DE LA SITUACIÓN PROBLEMÁTICA

Debido al número creciente de Personas con discapacidad visual, ha surgido la necesidad de herramientas que apoyen a Personas con discapacidad visual en labores cotidianas para reconocer objetos de uso diario, de manera rápida sin necesidad de exponerse a peligros, por ejemplo, cuando una Persona necesita peinarse debe buscar su peine en algún lugar de su hogar, para esto se hace muy complicado si no se sabe exactamente en qué lugar o zona se encuentra para poder peinarse. Es por ello que el etiquetado de imágenes tiene que ser la tarea de asignar etiquetas amigables para el ser humano a una imagen para que las etiquetas semánticas puedan reflejar mejor el contenido de la imagen y, por lo tanto, pueden ayudar a las Personas con discapacidad visual puedan acceder mejor a esa información y poder encontrar el objeto que necesitan. Pero esto no sucede en la actualidad, por lo general se tiene imágenes mal etiquetadas, es decir, la imagen tiene etiquetas que no corresponden a los objetos que tiene, lo que hace que las búsquedas de estas imágenes sean erróneas.

Otro caso es que la imagen tenga etiquetas incompletas, es decir, tenga las etiquetas de objetos que, si están en la imagen, pero ausencia de otras etiquetas que también posea.

Para lograr alternativas efectivas de solución en el etiquetado de imágenes es necesario evaluar en diferentes modelos de aprendizaje que logren captar las características y reconocer los patrones que poseen estas imágenes para lograr identificar las etiquetas correctas que poseen las imágenes. En este campo del aprendizaje profundo existen diferentes algoritmos como redes neuronales, redes neuronales convolucionales (CNN), Robust Principle Component Analysis+ CNN, Convolucional Auto Encoder + CNN, Noise Adaption layer + CNN y Noise Robust layer + CNN los cuales tienes diferentes métricas de rendimiento, pero pueden ser útiles para solucionar el problema del etiquetado de imágenes.

Otro problema que surge es que para utilizar estos algoritmos es necesario muchas imágenes, para lo cual se tiene que seleccionar una base de datos de imágenes que sea diversa en sus etiquetas, es decir, que posea muchas Gatoegorías de etiquetas e imágenes por cada etiqueta.

1.3 FORMULACIÓN DEL PROBLEMA

La discapacidad visual cuya dificultad de reconocer objetos y poder ubicarlos es un problema crítico. Por ello, es necesario diseñar y evaluar un método de etiquetado que ayude a las Personas con discapacidad visual a reconocer estos objetos y ubicarlos a través de imágenes.

¿Cómo diseñar un modelo efectivo para reconocer objetos automáticos que apoye a las Personas con discapacidad visual?

1.4 OBJETIVOS

1.4.1 Objetivo general

Diseñar un método efectivo de reconocimiento de objetos automático a través del etiquetado de imágenes, utilizando algoritmos de redes neuronales convolucionales, que contribuya como una herramienta de apoyo para Personas con discapacidad visual, cuya precisión media (AP o MAP) y puntuación F (F-Score) sea mayor al 80 %.

1.4.2 Objetivos específicos

- Pre procesar de imágenes para entrenamiento y evaluación.
- Diseñar y construir un modelo de solución para el etiquetado de imágenes automático. Probar el modelo de solución
- Determinar la combinación de parámetros que maximice la precisión media (AP o MAP) y la puntuación F(F-Score).

1.5 JUSTIFICACIÓN Y DELIMITACIÓN DE LA INVESTIGACIÓN

1.5.1 Justificación

En el Perú un tema de vital importancia es la necesidad que tienen las personas con discapacidad visual, a su vez las limitaciones que tienen para realizar actividades básicas, que para personas con ninguna discapacidad son tareas fáciles; es por esto que esta tesis se enfoca en este grupo humano que está en crecimiento en gran medida en la población mundial.

Por ende, es esencial desarrollar herramientas que alienten a las personas con discapacidad visual a tener una mejor calidad de vida y poder socializar de mejor manera, así como ofrecerles asistentes tecnológicos que estén a su alcance a bajo costo y que se pueda acceder sin necesidad de tener una gran cantidad de recursos.

1.5.2 Limitaciones

La presente tesis tiene como limitaciones el entrenamiento para 10 etiquetas de imágenes muy usadas, de las cuales se tiene muchas bases de datos en internet, a su vez se entrenará el modelo y se presentará un prototipo en aplicativo Android.

CAPÍTULO II FUNDAMENTO TEÓRICO

2.1 ANTECEDENTES

2.1.1 PAPER: Computer Vision-based Object Recognition for the Visually Impaired Using Visual Tags

Según los autores Rabia Jafri, Syed Abid Ali, y Hamid R. Arabnia Rconocer objetos genéricos en el entorno o en el medio ambiente es un gran desafío para las Personas con discapacidad visual. Sin embargo, en los últimos años, han surgido varias estrategias basadas en la visión por computadora, para esta tarea que utilizan etiquetas visuales pegadas a los objetos para su identificación.

Estos enfoques se distinguen por su dependencia de componentes comerciales y móviles, tecnologías que los hacen rentables, portátiles, intuitivos y por lo tanto, soluciones convincentes a un problema urgente.

En esta investigación se describe el estado de arte de las herramientas y soluciones tecnológicas que se han ideado, específicamente con el etiquetado de objetos, usando etiquetas físicas con reconocimiento RFID, o con un código de barras poder identificar que objeto es el que se tiene también el código QR que pueden tener, así como se describe en el Cuadro 1.

Para esto se hace uso de otras herramientas como dispositivos móviles que puedan tener al alcance fácil y sea transportable, Las herramientas que se describen son para reconocer estas etiquetas.

Approach	Type of tag	Input device	Output device	User interface to enter queries/ preferences
Badge3D [28]	1D barcode with black rectangular boundary	Head-mounted video camera	Headphones	Microphone
ShopMobile [32, 33]	MSI and UPC barcodes	Smartphone camera	Wireless over-the-ear headpiece	Smartphone touchscreen
Trinetra [5]	UPC barcodes	Barcode scanning pencil	Bluetooth headset	Smartphone touchscreen
Gude et al. [29]	Semacodes	Head-mounted and cane-mounted video cameras	Cane-mounted tactile Braille device	-
Al-Khalifa [37]	QR codes	QR reader-equipped Mobile phone camera	Mobile phone speakers	-
LookTel [11]	1.5" or 3" round, re-stickable vinyl stickers with printed images	Smart phone camera	Open ear, sports-designed headset	Smartphone touchscreen
Tekin et al. [38]	UPC-A barcodes	Mobile phone video camera	Mobile phone speakers	-
TalkingTag™ LV (Low Vision) [40]	2D barcode	iPhone camera	iPhone speakers	iPhone touchscreen

Cuadro N°1: Resumen de los enfoques de reconocimiento de objetos basados en la visión por computadora que usan etiquetas visuales

2.1.2 PAPER: Making Shopping Easy for People with Visual Impairment Using Mobile Assistive Technologies

Según los autores Mostafa Elgendy, Cecilia Sik-Lanyi, y Arpad Kelemen las Personas con discapacidad visual enfrentan diversas dificultades en sus actividades diarias en comparación con las Personas sin discapacidad visual. Se han realizado muchas investigaciones para encontrar soluciones inteligentes que utilizan dispositivos móviles para ayudar a las Personas con discapacidad visual a realizar tareas como ir de compras. Una de las tareas más difíciles para los investigadores es crear una solución que ofrezca una buena calidad de vida para las Personas con discapacidad visual. También es esencial desarrollar soluciones que alienten a las Personas con discapacidad visual a participar en la vida social. Este estudio proporciona una visión general de las diversas tecnologías que se han desarrollado en los últimos

años para ayudar a las Personas con discapacidad visual en las tareas de compra.

En la investigación nos muestra como la visión computacional es usada intensamente, y puede ser implementada en dispositivos móviles como lo son los teléfonos inteligentes. Una tarea de vital importancia para las Personas es realizar sus compras con facilidad, así que estas herramientas nos dan apoyo para poder apoyar a Personas con discapacidad visual, con el uso de tecnología de asistencia móvil llamada MAT por sus siglas en inglés.

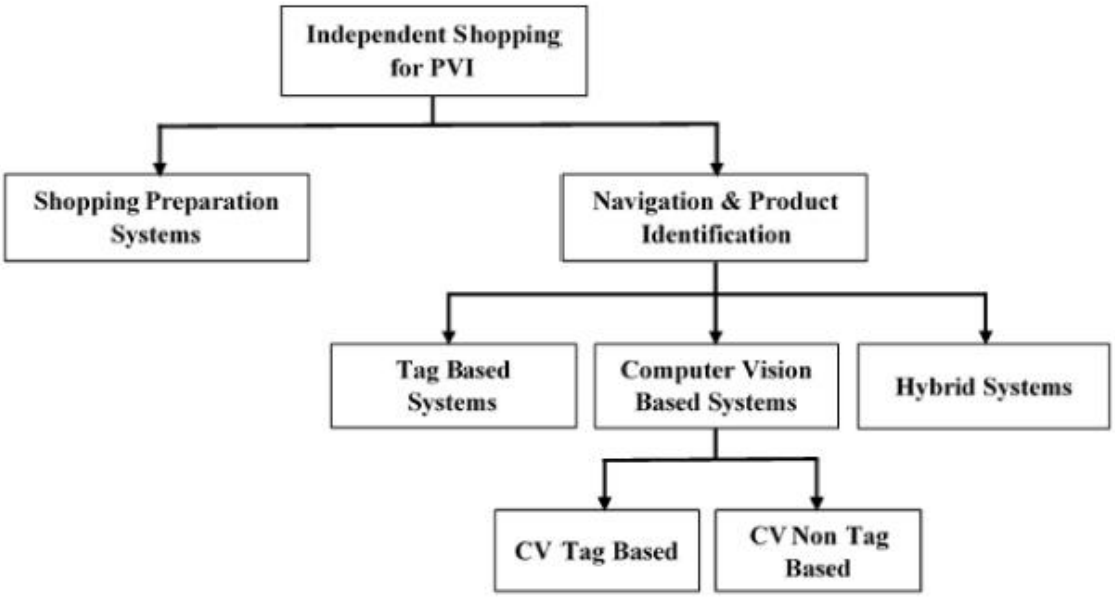


Figura N° 1: Soluciones de tecnología de asistencia móvil para las partes del proceso de compra para Personas con discapacidad visual

2.1.3 PAPER: An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends

Según los autores Alexy Bhowmick y Shyamanta M. Hazarika, la tecnología de asistencia para Personas con discapacidad visual y Personas ciegas es un campo de investigación que está aumentando prominencia debido a una explosión de nuevo interés en disciplinas de inteligencia artificial y visión computacional. El campo tiene un fuerte impacto social en nuestras poblaciones cada vez mayores y ciegas.

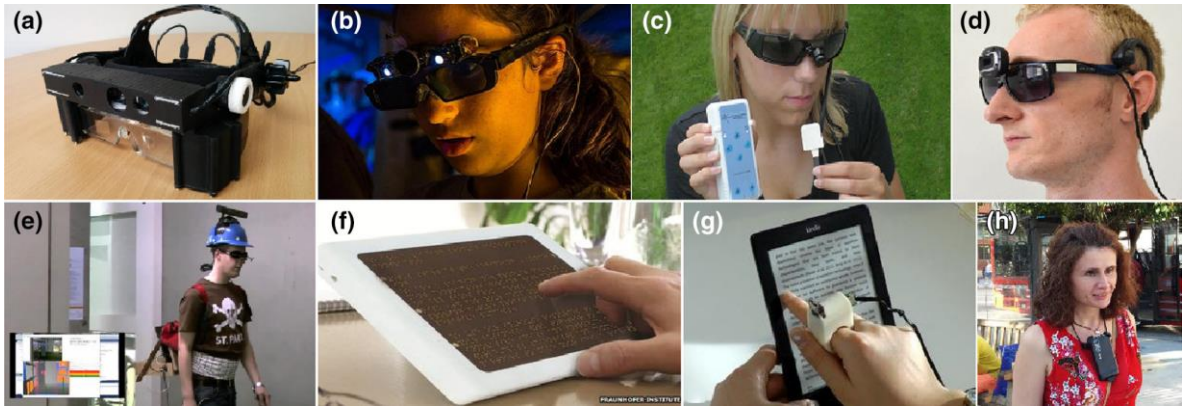


Figura 2: Ejemplos de tecnología de asistencia de investigaciones actuales para Personas con discapacidad visual y ciegas

2.1.4 PAPER: Advances in deep learning approaches for image tagging

Según los autores Jianlong Fu y Yong Rui la llegada de los dispositivos móviles y los servicios de medios en la nube ha llevado al crecimiento sin precedentes de las colecciones de fotos Personales. Uno de los problemas fundamentales en la gestión del creciente número de fotos es el etiquetado automático de imágenes. El etiquetado de imágenes es la tarea de asignar etiquetas amigables a los humanos a una imagen para que las etiquetas

semánticas puedan reflejar mejor el contenido de la imagen y por lo tanto, puede ayudar a los usuarios a acceder mejor a esa imagen. La calidad del etiquetado de imágenes depende de la calidad del modelado de conceptos que crea un mapeo desde los conceptos hasta las imágenes visuales. Si bien en la última década se han logrado avances signifiGatoivos en el etiquetado de imágenes, los enfoques anteriores solo pueden lograr un éxito limitado debido a la capacidad limitada de representación de conceptos de las características diseñadas a mano. Se avanzan más, ya que se han desarrollado algoritmos de aprendizaje profundo eficientes y efectivos. En este trabajo lo que describen es Gatoegorizar y evaluar diferentes enfoques de etiquetado de imágenes basados en técnicas de aprendizaje profundo. También discuten los problemas y aplicaciones relevantes al etiquetado de imágenes, incluida la recopilación de datos, las métricas de evaluación y los sistemas comerciales existentes. Concluimos las ventajas
de diferentes paradigmas de etiquetado de imágenes y proponer varias direcciones de investigación prometedoras para trabajos futuros.

En esta investigación se comparan los métodos de etiquetados de imágenes entre ellos un método muy eficiente las redes neuronales convolucionales que es el método aplicado en esta investigación para poder reconocer objetos automáticamente. En el Cuadro 2 se describen los tipos de métodos analizados en esta investigación.

Año	Método	Titulo	
2012	CNN	Imagenet classifiGatoion with deep convolutional neural networks	**Se usa el redes neuronales convolucionales para**
2011	RPCA+CNN	Segmentation of microcalcifiGatoion in X-ray mammo-grams using entropy thresholding	**Se usa redes neuronales convolucionales para Segmentación de microcalcificación en mamografías de rayos X**
2012	CAE+CNN	Hierarchical Face Parsing via Deep Learning	**Se usa aprendizaje profundo para el análisis jerárquico de rostros.**
2014	NA+CNN	Learning from noisy labels with deep neural networks	**Uso de redes neuronales convolucionales para el etiquetado de imágenes.**
2015	NR + CNN	**Relaxing from Vocabulary: Robust Weakly-Supervised Deep Learning for Vocabulary-Free Image Tagging**	**Aprendizaje profundo supervisado para etiquetado de imágenes sin vocabulario**

Cuadro 2: Métodos e investigaciones evaluados

2.1.5 TESIS: Estudio de métodos para identificar signos de retinopatía diabética en Imágenes de fondo del ojo

La autora Selene Montes Fuentes señala que la detección de la retinopatía diabética es una actividad que consume tiempo y generalmente se realiza de forma manual, debido a que se requiere de un médico capacitado para evaluar imágenes a color de la retina. Un problema común es la insuficiencia de médicos entrenados y de infraestructura necesaria en regiones donde se

presenta esta enfermedad. Si el número de Personas con diabetes aumenta, entonces será escasa tanto la infraestructura como el Personal para la prevención de la ceguera causada por la RD.

De lo anteriormente expuesto surge la necesidad de un procedimiento automatizado y completo, desarrollado mediante técnicas de clasificación de imágenes usando el aprendizaje automático y las redes convolucionales.

Por lo que propone una metodología para clasificar imágenes de fondo del ojo en las diferentes etapas de la retinopatía diabética y obtener un procedimiento eficaz que permita el diagnóstico de la enfermedad.

De esta tesis se aprecia que el uso del método de redes neuronales convolucionales es un método efectivo para reconocer patrones e imágenes y así poder identificar y clasificar correctamente con un grado de certeza muy alto, y de manera automática lo que permite su fácil uso y mejora en el tiempo.

2.1.6 TESIS: Desarrollo de una red neuronal convolucional para el procesamiento de imágenes placentarias

Según Omar Emilio Contreras Zaragoza el Deep learning es una técnica que ha ido creciendo en el análisis de datos. El aprendizaje profundo es una mejora de las redes neuronales artificiales, que consta de más capas que permiten mayores niveles de abstracción y predicciones mejoradas a partir de datos. Hasta la fecha ha emergido como la principal herramienta de aprendizaje automático en los dominios de visión computarizada

En particular las redes neuronales convolucionales (CNNs) han probado ser poderosas herramientas para una amplia gama de tareas de visión computarizada.

Deep CNNs automáticamente aprenden abstracciones de nivel medio y de alto nivel obtenidas a partir de datos brutos. Los resultados recientes indican que las características extraídas son extremadamente efectivas en el reconocimiento y localización de objetos en imágenes naturales. Los grupos de imágenes médicas en todo el mundo están ingresando al campo rápidamente y están aplicando CNN y otras metodologías de Deep learning para una amplia variedad de aplicaciones.

En el trabajo se le da una nueva aplicación a una red neuronal convolucional profunda en la que se segmentan venas y arterias de una imagen placentaria inyectada con colorante para facilitar su análisis y poder obtener más información sobre el síndrome de transfusión gemelo-gemelo, se presenta la metodología utilizada con la que se llegó a una arquitectura novedosa y a través de la cual se llega a muy buenos resultados que también se exponen numéricamente y visualmente.

2.2 MARCO TEÓRICO CONCEPTUAL

2.2.1 Técnicas de aprendizaje profundo

Las técnicas de aprendizaje profundo se refieren a una clase de técnicas de aprendizaje automático, donde se explotan muchas capas de etapas de procesamiento de información en arquitecturas jerárquicas para el aprendizaje de funciones sin supervisión y para el análisis/clasificación de patrones supervisados. El procedimiento principal del aprendizaje profundo es calcular características o representaciones jerárquicas de los datos observados, donde las características o facto- res de nivel superior se definen a partir de la estructura de datos de nivel inferior. Las técnicas de aprendizaje profundo generalmente se pueden dividir en tres dimensiones, es decir, arquitecturas

profundas generativas, arquitecturas profundas discriminativas y arquitecturas profundas hibridas.

2.2.2 Redes Neuronales

Las redes neuronales son un modelo computacional que está formadas por un conjunto de neuronas artificiales interconectadas. Las neuronas de la red se encuentran distribuidas en diferentes capas de neuronas, de manera que las neuronas de una capa están conectadas con las neuronas de la capa siguiente, a las que pueden enviar información. La arquitectura más usada en la actualidad de una red neuronal (como la presentada en la figura 2.1) consistiría en:

1. Una primera capa de entradas, que recibe información del exterior.
2. Una serie de capas ocultas (intermedias), encargadas de realizar el trabajo de la red.
3. Una capa de salidas, que proporciona el resultado del trabajo de la red al exterior.

El número de capas intermedias y el número de neuronas de cada capa dependerá del tipo de aplicación al que se vaya a destinar la red neuronal. Cada neurona de la red es una unidad de procesamiento de información; es decir, recibe información a través de las conexiones con las neuronas de la capa anterior, procesa la información, y emite el resultado a través de sus conexiones con las neuronas de la capa siguiente, siempre y cuando dicho resultado supere un valor umbral que es determinado con una función de activación.

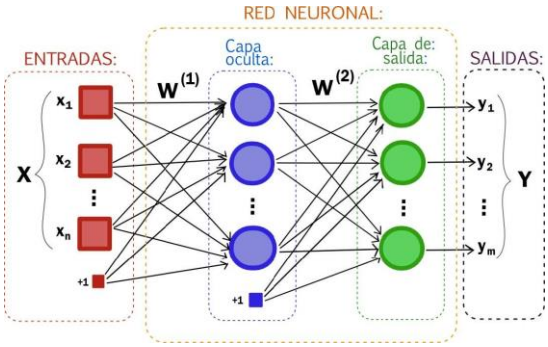


Figura 2.1: Red neuronal artificial

En una red neuronal ya entrenada, las conexiones entre neuronas tienen un determinado peso (“peso sináptico”).

En la figura 2.2 se observa una función F que es igual a la cual es evaluada por una función de activación y de acuerdo a esa evaluación tendrá un valor de salida hacia la siguiente neurona.

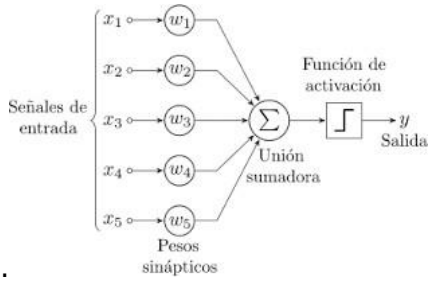


Figura 2.2: Función de Activación

Al finalizar el procesamiento se obtendrá´ un valor de y el cual junto con el valor de y real que se quiere obtener de la red neuronal, serán evaluados en una función de coste para hallar el error generado. Una vez hallado este valor se origina un proceso hacia atrás cuyo objetivo es ajustar los pesos asignados para minimizar la función de coste.

2.2.2.1 Implementación de una red neuronal

Pasos para la implementación de una red neuronal (ver figura 2.3)

1. Inicializar aleatoriamente los pesos a números pequeños cerca de 0.
2. Ingresar la primera observación de su conjunto de datos en la capa de entrada, cada característica en un nodo de entrada.
3. Propagación hacia delante: de izquierda a derecha, las neuronas se activan de forma tal que el impacto de la activación de cada neurona esta´ limitado por los pesos. Propaga las activaciones hasta obtener el resultado predicho.
4. Se procede a comparar el resultado predicho con el resultado real. A continuación se mide error generado.
5. Back - Propagation: de derecha a izquierda, el error se propaga hacia atrás. Actualice las ponderaciones de acuerdo del error. La tasa de aprendizaje decide por cuanto actualizamos los pesos.
6. Repetir los pasos 1 al 5 y se actualizan los pesos después de cada observación

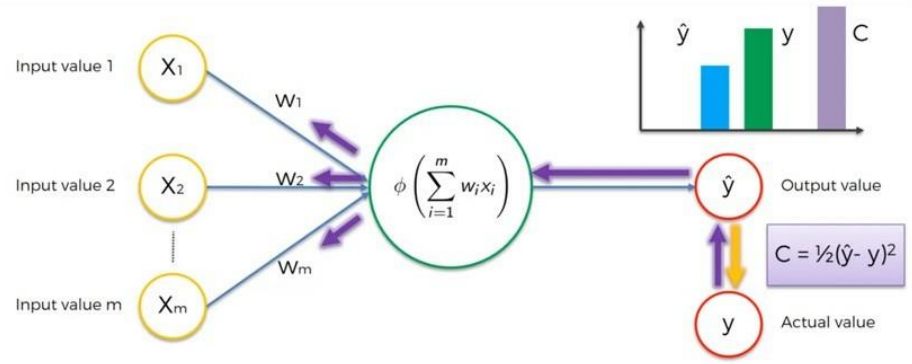


Figura 2.3: Back-Propagation y función de coste

2.2.3 Redes Neuronales Convolucionales

Las redes neuronales convolucionales son muy similares a las redes neuronales ordinarias; se componen de neuronas que tienen pesos y sesgos que pueden aprender. Cada neurona recibe algunas entradas, realiza un producto escalar y luego aplica una función de activación. Al igual que en el perceptrón multicapa también vamos a tener una función de pérdida o costo sobre la última capa, la cual estará totalmente conectada. Lo que diferencia a las redes neuronales convolucionales es que suponen explícitamente que las entradas son imágenes, lo que nos permite codificar ciertas propiedades en la arquitectura; permitiendo ganar en eficiencia y reducir la cantidad de parámetros en la red.

Las redes neuronales convolucionales trabajan modelando de forma consecutiva pequeñas piezas de información, y luego combinando esta información en las capas más profundas de la red. Una manera de entenderlas es que la primera capa intentara´ detectar los bordes y establecer patrones de detección de bordes. Luego, las capas posteriores trataran de combinarlos en formas más simples y, finalmente, en patrones de las diferentes posiciones de los objetos, iluminación, escalas, etc. Las capas finales intentarán hacer coincidir una imagen de entrada con todos los patrones y arribar a una predicción final como una suma ponderada de todos ellos. De esta forma las redes neuronales convolucionales son capaces de modelar complejas variaciones y comportamientos dando predicciones bastantes precisas.

2.2.3.1 Estructura de las Redes Neuronales Convolucionales

En general, las redes neuronales convolucionales van a estar construidas con una estructura que contendrá 3 tipos distintos de capas:

1. Una capa convolucional, que es la que le da el nombre a la red.
2. Una capa de reducción o de pooling, la cual va a reducir la cantidad de parámetros al que- darse con las características más comunes.
3. Una capa clasificadora totalmente conectada, la cual nos va dar el resultado final de la red.

2.2.3.2 Capa convolucional

La operación de convolución recibe como entrada o input la imagen y luego aplica sobre ella un filtro o kernel que devuelve un mapa de las características de la imagen original, de esta forma logramos reducir el tamaño de los parámetros, como se observa en la figura 2.4. Además por cada operación con los diferentes filtros se obtiene el mismo número de mapas de características (figura 2.5)

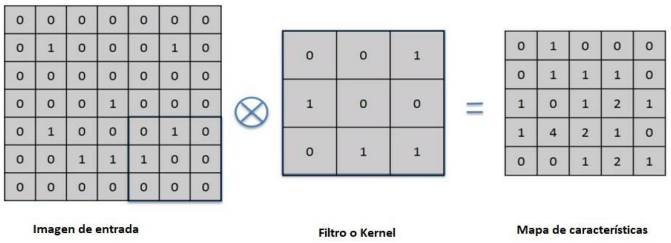


Figura 2.4: Operación de convolución

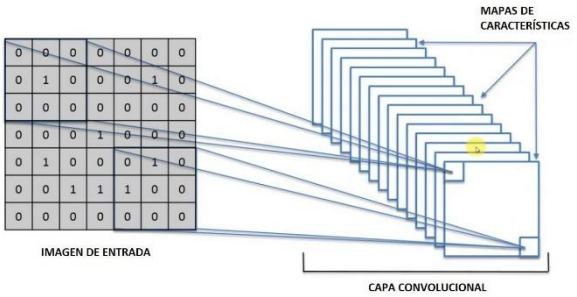


Figura 2.5: Obtención de mapas de características

2.2.3.3 Capa de reducción o pooling

La capa de reducción o pooling se coloca generalmente después de la capa convolucional. Su utilidad principal radica en la reducción de las dimensiones espaciales (ancho x alto) del volumen de entrada para la siguiente capa convolucional. No afecta a la dimensión de profundidad del volumen. La operación realizada por esta capa también se llama reducción de muestreo, ya que la reducción de tamaño conduce también a la pérdida de información. Sin embargo, una pérdida de este tipo puede ser beneficioso para la red por dos razones:

1. La disminución en el tamaño conduce a una menor sobrecarga de cálculo para las próximas capas de la red.
2. Reducir el sobreajuste

La operación que se suele utilizar en esta capa es max-pooling, que divide a la imagen de entrada en un conjunto de rectángulos y, respecto de cada subregión, se va quedando con el máximo valor, como se observa en la figura 2.6

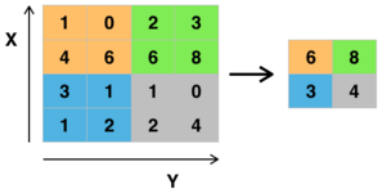


Figura 2.6: Operación de max-pooling

2.2.3.4 Métricas de rendimiento

Para evaluar el desempeño de diferentes enfoques de etiquetado de imagen / refinamiento de etiqueta, se han propuesto varias medidas de evaluación en la comunidad de investigación, incluyendo puntuación F, precisión media (AP). Además se utilizara´ comparar los valores obtenidos al evaluar el modelo.

Matriz de confusión común para dos clases ver cuadro 2.1.

		Valor Predicho	
		Positivos	*Negativos*
Valor real	*Positivos*	Verdaderos positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos positivos (FP)	Verdaderos negativos (VN)

Cuadro 2.1: Matriz de confusión simple

Donde:

- VP es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo.
- VN es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo.

- FN es la cantidad de positivos que fueron clasificados incorrectamente como negativos. FP es la cantidad de negativos que fueron clasificados incorrectamente como positivos.
- De la matriz de confusión se derivan métricas de medidas de eficiencia las cuales son puntación F y precisión media que se explicaran a continuación.
- Para calcular la puntuación F se necesitan los valores de precisión y recuperación del modelo, para la precisión media se necesita los valores de precisión de predicción de cada etiqueta o clase.
- Ecuación de precisión (ver ecuación 2.1) y recuperación (ver ecuación 2.2) simple para calcular la eficiencia en una predicción de dos clases, tomando en cuenta como elementos relevantes los positivos.

De la matriz de confusión se derivan métricas de medidas de eficiencia las cuales son puntación F y precisión media que se explicarán a continuación.

Para calcular la puntuación F se necesitan los valores de precisión y recuperación del modelo, para la precisión media se necesita los valores de precisión de predicción de cada etiqueta o clase.

Ecuación de precisión (ver ecuación 2.1) y recuperación (ver ecuación 2.2) simple para calcular la eficiencia en una predicción de dos clases, tomando en cuenta como elementos relevantes los positivos.

$$Precisión\ Positivos = PP = \frac{VP}{VP + FP} \qquad (2.1)$$

$$Recuperación\ Positivos = RP = \frac{VP}{VP + FN} \qquad (2.2)$$

La ecuación para calcular la puntuación F (F Score) tomando en cuenta como elementos relevantes los positivos es la ecuación 2.3.

$$Puntuación\ F\ positivos = PFP = \frac{2(PP)(RP)}{PP + RP} \qquad (2.3)$$

Entonces la ecuación de precisión (ver ecuación 2.4) y recuperación (ver ecuación 2.5) simple para calcular la eficiencia en una predicción de dos clases, tomando en cuenta como elementos relevantes los negativos, cambia los valores a evaluar en las ecuaciones.

$$Precisión\ Negativos = PN = \frac{VN}{VN + FN} \qquad (2.4)$$

$$Recuperación\ Negativos = RN = \frac{VN}{VN + FP} \qquad (2.5)$$

La ecuación para calcular la puntuación F (F Score) tomando en cuenta como elementos relevantes los negativos es la ecuación 2.6

$$Puntuación\ F\ negativos = PFN = \frac{2(PN)(RN)}{PN + RN} \qquad (2.6)$$

La ecuación 2.7 calcula la precisión media del modelo.

$$Precisión\ Media = PFN = \frac{PP + PN}{2} \qquad (2.7)$$

Por lo tanto la puntuación F (F Score) es aplicada para cada etiqueta o clase, entonces se calcula la puntuación F media que es un valor de eficiencia para el modelo según la ecuación 2.8.

$$Precisión\ F\ Media = PFN = \frac{PFP + PFN}{2} \qquad (2.8)$$

En la presente tesis se usa un número mayor de etiquetas para lo cual la matriz de confusión, la precisión y la recuperación cambian de ecuación por lo tanto también lo hacen la puntuación y la precisión media.

Para un número mayor de etiquetas la matriz confusión (ver cuadro 2.2) sería distinta a la matriz de confusión simple (ver cuadro 2.1).

		Valor Predicho			
		Etiqueta 1	Etiqueta 2	...	*Etiqueta N*
Valor real	Etiqueta 1	Q_{11}	Q_{12}	...	Q_{1N}
	Etiqueta 2	Q_{21}	Q_{22}	...	Q_{2N}
	...	...	...	...	...
	Etiqueta N	Q_{N1}	Q_{N2}	...	Q_{NN}

Cuadro 2.2: Matriz de confusión

Entonces la ecuación de precisión (ver ecuación 2.1) y recuperación (ver ecuación 2.2) cambian para múltiples etiquetas como se aprecia en las ecuaciones 2.9 y 2.10 para la etiqueta i.

$$Precisión_i = P_i = \frac{Q_{ii}}{\sum_{i=1}^{i=N}(Q_{ji})} \qquad (2.9)$$

$$Precisión_i = R_i = \frac{Q_{ii}}{\sum_{i=1}^{i=N}(Q_{ij})} \qquad (2.10)$$

La ecuación para calcular la puntuación F (F Score) tomando en cuenta como elementos relevantes a la etiqueta i es la ecuación 2.11

$$Puntuación\ F_i = PF_i = \frac{2(P_i)(R_i)}{P_i + R_i} \qquad (2.11)$$

La ecuación 2.12 calcula la precisión media del modelo, donde N es el número de etiquetas.

$$Precisión\ media = \frac{\sum_{i=1}^{i=N}(P_i)}{N} \qquad (2.12)$$

Por lo tanto la puntuación F (F Score) es aplicada para cada etiqueta o clase, entonces se calcula la puntuación F media que es un valor de eficiencia para el modelo según la ecuación 2.13.

$$Precisión\ F\ media = \frac{\sum_{i=1}^{i=N}(PF_i)}{N} \qquad (2.13)$$

2.3 MARCO TEÓRICO INSTRUMENTAL

Se selecciona los siguientes enfoques típicos de etiquetado de imágenes basados en modelos para comparar [3]:

- **CNN:** Convolutional Neural Network adopta la arquitectura de red del estado de la técnica con varias capas convolucionales y capas conectadas totalmente.

- **RPCA + CNN:** Robust Principle Component Analysis + CNN quita primero las muestras con grandes errores de reconstrucción por RPCA, y lleva a cabo la formación CNN sobre las muestras limpiadas.

- **CAE + CNN:** Convolucional Auto Encoder + CNN propone reducir el efecto del ruido en el entrenamiento de la CNN mediante la estrategia de ajuste previo y ajuste de la capa.

- **NA + CNN:** Noise Adaption layer + CNN propone agregar una capa adicional de adaptación de ruido de abajo hacia arriba en la arquitectura CNN tradicional para la eliminación de ruido.

- **NR + CNN:** Noise Robust layer + CNN diseña una función objetivo para asignar esas muestras ruidosas, pocas y diferentes, con autoridades de bajo nivel en el entrenamiento, y así se puede reducir el efecto del ruido.

Cuadro 2.3: Trabajos de investigación según el método usado.

Autores	Año	Método	Titulo
Krizhevsky, A.; Sutskever, I.; Hinton, G.E. [6]	2012	CNN	**Imagenet classifiGatoion with deep convolutional neural networks**
Candes, E.J.; Li, X.; Ma, Y.; Wright, J. [1]	2011	RPCA+CNN	**Segmentation of microcalcifiGatoion in X-ray mammo-grams using entropy thresholding**
Luo, P.; Wang, X.; Tang, X. [7]	2012	CAE+CNN	**Hierarchical Face Parsing via Deep Learning**
Sukhbaatar, S.; Fergus, R. [8]	2014	NA+CNN	**Learning from noisy labels with deep neural networks**
Fu, J.; Wu, Y.; Mei, T.; Wang, J.; Lu,H.; Rui, Y. [4]	2015	NR + CNN	**Relaxing from Vocabulary: Robust Weakly-Supervised Deep Learning for Vocabulary-Free Image Tagging**

CAPÍTULO III DESCRIPCIÓN DE DATOS

3.1 FUENTE DE DATOS

Para realizar el objetivo del presente trabajo se utilizó´ la base de datos de imágenes llamada ImageNet, cuyo conjunto de datos de imágenes están organizados de acuerdo a la jerarquía de WordNet. Cada concepto signifiGatoivo en WordNet, posiblemente descrito por varias palabras o frases de palabras, se denomina conjunto de sinónimos o ”synset”. Hay más de 100,000 synsets en WordNet, la mayoría de ellos son sustantivos (80,000+).

3.1.1 Obtención de la base de datos

El proceso para la obtención de datos es el siguiente (Figura 3.1)

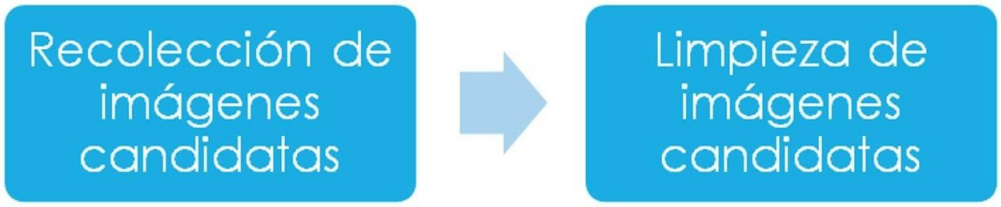


Figura 3.1: Proceso de obtención de la base de datos

3.1.1.1 Recopilación de imágenes candidatas

Como se describe en [5] la primera etapa de la construcción de ImageNet implica la recopilación de imágenes candidatas para cada synset. En esta etapa se recopilaron imágenes candidatas de internet mediante la consulta de varios motores de búsqueda de imágenes para cada grupo de sinónimos o synset de WordNet [2]. Para la obtención de una mayor cantidad de imágenes

posible, se expandió´ el conjunto de consultas agregando las consultas con la palabra del synset principal. Por ejemplo, al realizar la consulta”whippet”, que de acuerdo a la definición descrita en WordNet es ”perro pequeño y delgado de tipo galgo desarrollado en Inglaterra. Entonces las siguientes consultas serán ”whippet Perro ”whippet greyhound”. Para la ampliación y diversificación aún más del grupo de candidatos se traduce las consultas en otros idiomas como: chino mandarín, español, holandés e italiano.

3.1.1.2 Limpieza de las imágenes candidatas

En esta etapa se realizó la limpieza de las imágenes candidatas obtenidas en el paso anterior con el objetivo de recopilar un conjunto de datos con alta precisión. Esto se logró mediante el uso del servicio Amazon Mechanical Turk (AMT), una plataforma en línea que facilita a las Personas y a las empresas subcontratar sus procesos y trabajos para que usuarios las completen y se les paguen. En cada una de las tareas de etiquetado, se presenta al usuario un conjunto de imágenes candidatas y la definición del conjunto objetivo. A continuación, se les pide a los usuarios verificar si cada imagen contiene objetos del synset indicado. Además, se le solicita al usuario seleccionar imágenes independientemente de la cantidad de objetos y el desorden de la escena para garantizar la diversidad.

Para cada synset, primero se muestrea aleatoriamente un subconjunto inicial de imágenes. Al menos 10 usuarios deben votar en cada una de estas imágenes. Luego se este procedimiento se obtiene una tabla de puntuación de confianza (Figura 3.2) que indica la probabilidad de que una imagen sea

una buena imagen dados los votos de los usuarios. Para algunos synsets, los usuarios no logran obtener un voto mayoritario para ninguna imagen, lo que indica que el synset no se puede ilustrar fácilmente con imágenes.

User 1	Y	Y	Y
User 2	N	Y	Y
User 3	N	Y	Y
User 4	Y	N	Y
User 5	Y	Y	Y
User 6	N	N	Y

#Y	# N	Conf Cat	Conf BCat
0	1	0.07	0.23
1	0	0.85	0.69
1	1	0.46	0.49
2	0	0.97	0.83
0	2	0.02	0.12
3	0	0.99	0.90
2	1	0.85	0.68

Figura 3.2: Izquierda: ¿Hay un gato birmano en las imágenes? Seis usuarios muestreados al azar tienen diferentes respuestas. A la derecha: la tabla de puntuación de confianza para "Gato birmano". Se necesitan más votos para alcanzar el mismo grado de confianza para las imágenes del "gato birmano".

3.2 ESTRUCTURA DE LOS DATOS

La base de datos de imágenes llamada ImageNet se basa en la estructura jerárquica proporcionada por WordNet. Cada concepto signifiGatoivo en WordNet, posiblemente descrito por varias palabras o frases de palabras, se denomina conjunto de sinónimos o ”synset”. Hay alrededor de 80000 sintetizadores nominales en WordNet. ImageNet, consta de 12 subárboles: mamíferos, aves, peces, reptiles, anfibios, vehículos, muebles, instrumentos musicales, formación geológica, herramientas, flores, frutas. Estos subárboles contienen 5247 synsets y 3.2 millones de imágenes.

La figura 3.3 muestra una instantánea de dos ramas de los subárboles de mamíferos y vehículos.

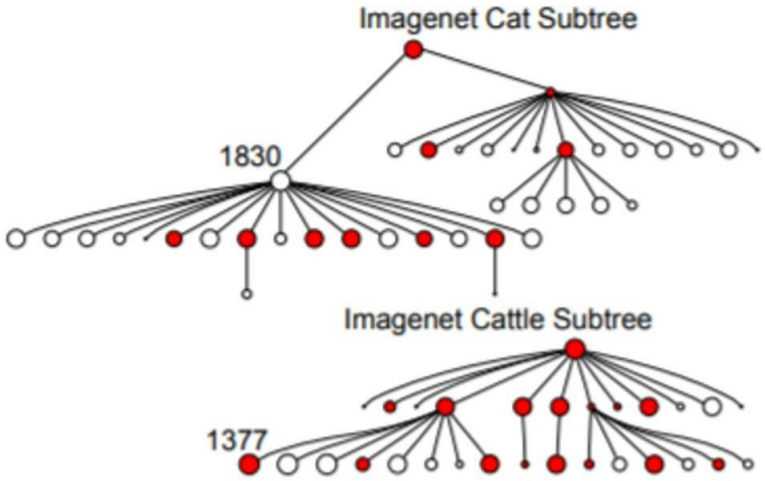


Figura 3.3: Subárboles de los synsets ”Gato ”Ganado” de ImageNet

3.3 DESCRIPCIÓN DE LOS DATOS

El conjunto de imágenes de ImageNet presenta las siguientes propiedades:

Escala: ImageNet tiene como objetivo proporcionar la cobertura más completa y diversa del mundo de la imagen. Los 12 subárboles actuales consisten en un total de 3.2 millones de imágenes con anotaciones limpias distribuidas en 5247 Gatoegorías. En promedio se recolectan más de 600 imágenes para cada synset.

Jerarquía: ImageNet organiza las diferentes clases de imágenes en una jerarquía semántica densamente poblada. El principal activo de WordNet reside en su estructura semántica, es decir, su ontología de conceptos. De manera similar a WordNet, los conjuntos de imágenes en ImageNet están

interconectados por varios tipos de relaciones, siendo la relación ¨IS-A"la más completa y útil.

Precisión: Imagenet ofrece un conjunto de datos limpio en todos los niveles de la jerarquía de WordNet.

Diversidad: ImageNet se construye con el objetivo de que los objetos en las imágenes deban tener aspectos, posiciones, puntos de vista y posturas variables, así como un desorden de fondo y oclusiones.

3.3.1 Estadística Univariada

Número total de synsets no vacíos: 21841

Número total de imágenes: 14 197 122

Número de imágenes con anotaciones en el cuadro delimitador: 1.034.908

Número de synsets con características SIFT: 1000

Número de imágenes con características SIFT: 1.2 millones

CAPÍTULO IV MODELO DE SOLUCIÓN

4.1 MODELO DE SOLUCIÓN

El modelo de solución planteado en la presente tesis esta´ basado en el modelo de redes neuronales convolucionales [6]. En el algoritmo utilizado, las variables de entrada son las imágenes a procesar y las mediciones descritas (ver capítulo descripción de los datos). El procesamiento de las imágenes

esta´ sujeto a valores múltiples de parámetros y métricas, que se utilizan para cada capa en la construcción del modelo, descrito en 3. Como resultado del algoritmo se obtiene las etiquetas de las imágenes de entrada, así como las medidas de eficiencia precisión media (AP o MAP) y la puntuación F(F-Score). El esquema general de la solución empleada se muestra en la figura 4.1:

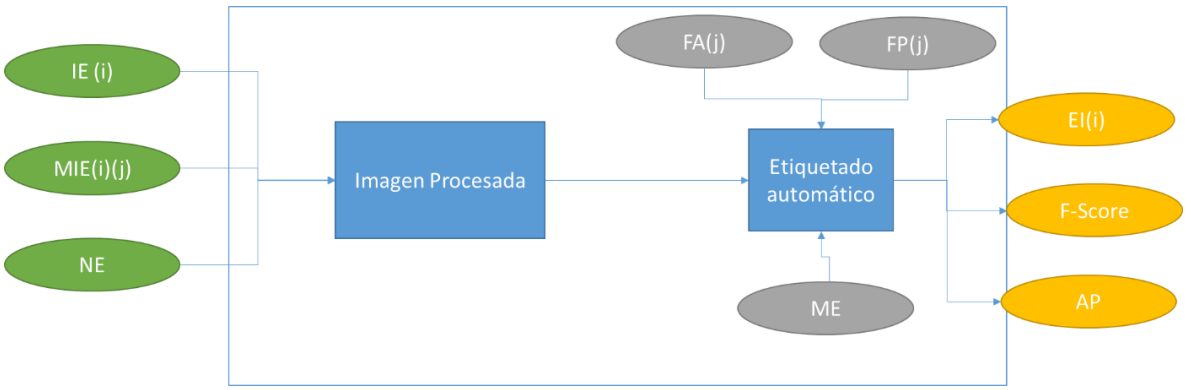


Figura 4.1: Modelo Solución

Las variables y parámetros que intervienen en el modelo de solución se describen en 4.1

Cuadro 4.1: Variables del modelo solución

Variable	**Descripción**
IE(i)	Imágenes de entrada por cada clase i
MIE(i)(j)	Métricas de cada imagen j de la clase i (peso Kb, tamaño píxeles)
NE	Es el número de etiquetas que tendrá el modelo solución

FA(j)	Es la función de activación que se usará para el entrenamiento, para cada capa (j).
FP(j)	Es la función de pérdida que se usará para el entrenamiento, para cada capa (j).
ME	Modelo entrenado
EI(i)	Etiqueta de imagen i
AP	Precisión media, métrica de eficiencia.
F-Score	Puntación F, métrica de eficiencia.

El esquema detallado de la solución empleada se muestra en la figura 4.2:

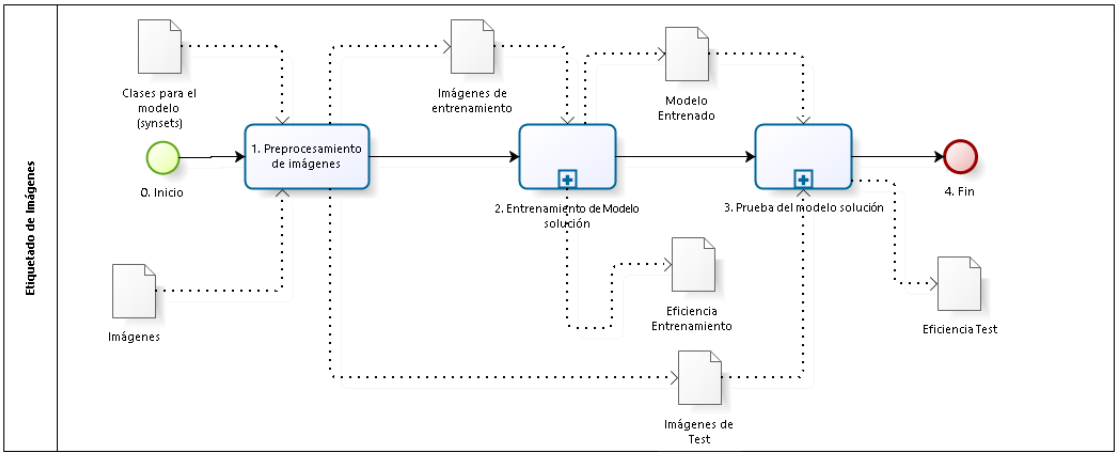


Figura 4.1: Procesos del Modelo Solución

El modelo multicapa se presenta en la figura 4.2:

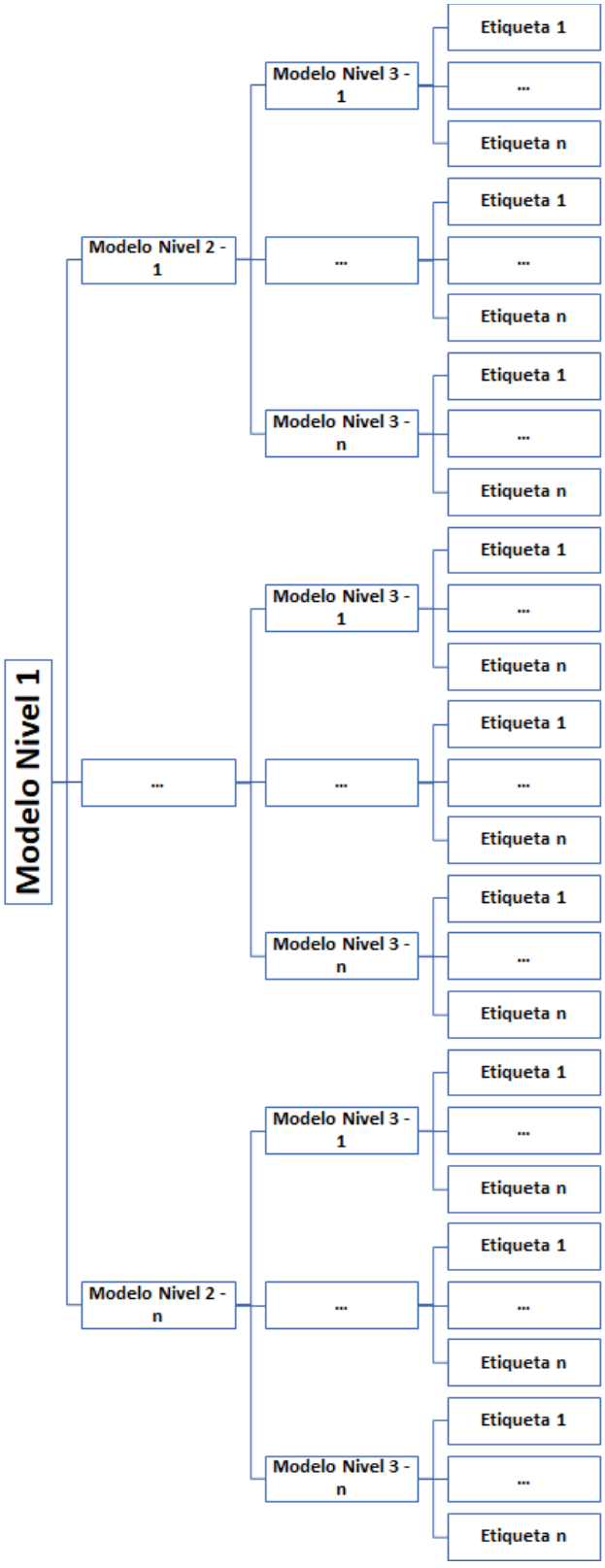


Figura 4.2: Modelo Multicapa

- En este modelo se tiene **78** etiquetas o clases
- Este modelo consiste en entrenar múltiples modelos, ya que cuando se tienen menor número de clases por etiqueta (en este caso **133** en promedio por clase), si fuera evaluado con un solo modelo tendría muy baja eficiencia, ya que las CNN necesitan una gran cantidad de muestras por clase para una correcta clasificación o predicción.
- La imagen muestra cómo funciona el modelo multicapa, en este caso solo hasta el nivel 3, hay 10 modelos.

En la figura 4.3 se muestra el modelo solución multicapa, que mejora la eficiencia a un solo modelo para 78 etiquetas.

Figura 4.3: Modelo Multicapa Solución

4.1.1 Pre procesamiento de imágenes

Este primer procedimiento homogeniza las imágenes de entrada, tanto en tamaño (alto y ancho) se redefine la cantidad de pixeles de alto y ancho, así como su tamaño; para luego realizar un Split o separación de imágenes para el entrenamiento y prueba del modelo, en el cual para las imágenes de entrenamiento se seleccionan el 75 % para el entrenamiento y el 25 % para la posterior evaluación del modelo. Este procedimiento tiene como entrada las imágenes etiquetadas, así como el número de etiquetas que se tienen que son 10 ver tabla 4.2:

Cuadro 4.2: Número de imágenes por etiqueta

Etiqueta	Número de imágenes
Ave	1308
Gato	1044
Perro	824
Pez	742
Flor	751
Comida	1067
Persona	844
Reptil	705
Árbol	60
Utensilio	894

El ancho y alto determinado para el pre procesamiento de imágenes es de 64 por 64 pixeles, imágenes a color (RGB) como se aprecia en la tabla 4.3:

Cuadro 4.3: Dimensión vs Pixeles vs Cantidad de valores por pixel

	Píxeles	Cantidad de valores por píxel
Alto	64	-
Ancho	64	-
Color	-	3

El Split para las imágenes de la tabla 4.3(Etiqueta Numero de imágenes) sería como en la tabla 4.4:

Cuadro 4.4: División de la cantidad de imágenes para el entrenamiento y prueba del modelo

Etiqueta	Número de imágenes	Imágenes de entrenamiento	Imágenes de prueba
Ave	1308	981	327
Gato	1044	783	261
Perro	824	618	206
Pez	742	557	186

Flor	751	563	188
Comida	1067	800	267
Persona	844	633	211
Reptil	705	529	176
Árbol	60	45	15
Utensilio	894	671	224

Para el modelo Multicapa se usa el mismo procedimiento como se aprecia en la tabla 4.5.

Cuadro 4.5: Número de imágenes por etiqueta – Modelo Multicapa

	Etiquetas	Número de imágenes	75 %	25%	Etiquetas Nivel 2	Etiquetas Nivel 3
1	Agua Mineral_Cielo	144	108	36	Cielo	
2	Agua Mineral_Otros	227	170	57	Otros	
3	Agua Mineral_San Carlos	89	67	22	San Carlos	
4	Agua Mineral_San Luis	144	108	36	San Luis	
5	Agua Mineral_San Mateo	119	89	30	San Mateo	
6	Café_Altomayo	76	57	19	Altomayo	
7	Café_Cafetal	71	53	18	Cafetal	
8	Café_Kirma	69	52	17	Kirma	
9	Café_Nescafé	69	52	17	Nescafé	
10	Café_Otros	223	167	56	Otros	

11	Gaseosa_7up	71	53	18	7up	
12	Gaseosa_Coca Cola	217	163	54	Coca Cola	
13	Gaseosa_Crush	66	50	17	Crush	
14	Gaseosa_Fanta	78	59	20	Fanta	
15	Gaseosa_Guaraná	81	61	20	Guaraná	
16	Gaseosa_Inka Kola	153	115	38	Inka Kola	
17	Gaseosa_KR	70	53	18	KR	
18	Gaseosa_Otros	156	117	39	Otros	
19	Gaseosa_Pepsi	138	104	35	Pepsi	
20	Gaseosa_Sprite	130	98	33	Sprite	
21	Jabón_Asepxia	222	167	56	Asepxia	
22	Jabón_Bolivar	80	60	20	Bolivar	
23	Jabón_Camay	219	164	55	Camay	
24	Jabón_Dove	295	221	74	Dove	
25	Jabón_Johnson	237	178	59	Johnson	
26	Jabón_Lux	295	221	74	Lux	
27	Jabón_Neko	75	56	19	Neko	
28	Jabón_Nivea	151	113	38	Nivea	
29	Jabón_Otros	129	97	32	Otros	
30	Jabón_Palmolive_Delicada Exfoliación	137	103	34	Palmolive_Delicada Exfoliación	Delicada Exfoliación
31	Jabón_Palmolive_Otros	219	164	55	Palmolive_Otros	Otros
32	Jabón_Palmolive_Sensación Humectante	151	113	38	Palmolive_Sensación Humectante	Sensación Humectante
33	Jabón_Palmolive_Suavidad Relajante	128	96	32	Palmolive_Suavidad Relajante	Suavidad Relajante
34	Jabón_Protex	146	110	37	Protex	
35	Otros_Árboles	157	118	39	Árboles	
36	Otros_Cama	175	131	44	Cama	
37	Otros_Cepillo	75	56	19	Cepillo	
38	Otros_Cubiertos	165	124	41	Cubiertos	

39	Otros_Fondos de Colores	139	104	35	Fondos de Colores	
40	Otros_Fondos de lugares dentro de casa	142	107	36	Fondos de lugares dentro de casa	
41	Otros_Gato	78	59	20	Gato	
42	Otros_Lapicero	83	62	21	Lapicero	
43	Otros_Laptop	55	41	14	Laptop	
44	Otros_Lata	78	59	20	Lata	
45	Otros_Mano	75	56	19	Mano	
46	Otros_Mesa	84	63	21	Mesa	
47	Otros_Mouse	62	47	16	Mouse	
48	Otros_Olla	81	61	20	Olla	
49	Otros_Otros animales	167	125	42	Otros animales	
50	Otros_Otros objetos	157	118	39	Otros objetos	
51	Otros_Pasto	77	58	19	Pasto	
52	Otros_Perro	29	22	7	Perro	
53	Otros_Persona	106	80	27	Persona	
54	Otros_Silla	163	122	41	Silla	
55	Otros_Sillón	157	118	39	Sillón	
56	Otros_Taza	85	64	21	Taza	
57	Otros_Vaso	89	67	22	Vaso	
58	Otros_Vegetación	95	71	24	Vegetación	
59	Pasta Dental_Aquafresh	143	107	36	Aquafresh	
60	Pasta Dental_Colgate	208	156	52	Colgate	
61	Pasta Dental_Kolynos	83	62	21	Kolynos	
62	Pasta Dental_Oral B	169	127	42	Oral B	
63	Pasta Dental_Otros	161	121	40	Otros	
64	Pasta Dental_Sensodyne	166	125	42	Sensodyne	
65	Shampoo_Dove	184	138	46	Dove	
66	Shampoo_Elvive	132	99	33	Elvive	
67	Shampoo_H&S	197	148	49	H&S	

68	Shampoo_Herbal Essences	112	84	28	Herbal Essences	
69	Shampoo_Johnson	119	89	30	Johnson	
70	Shampoo_Otros	189	142	47	Otros	
71	Shampoo_Palmolive	224	168	56	Palmolive	
72	Shampoo_Pantene_Fuerza y Reconstrucción	129	97	32	Pantene_Fuerza y Reconstrucción	Fuerza y Reconstrucción
73	Shampoo_Pantene_Hidratación Extrema	75	56	19	Pantene_Hidratación Extrema	Hidratación Extrema
74	Shampoo_Pantene_Otro	145	109	36	Pantene_Otro	Otro
75	Shampoo_Pantene_Restauración	135	101	34	Pantene_Restauración	Restauración
76	Shampoo_Pert Plus	145	109	36	Pert Plus	
77	Shampoo_Savital	135	101	34	Savital	
78	Shampoo_Sedal	143	107	36	Sedal	
	TOTAL	10443	7838	2618		
	PROMEDIO	134				

4.1.2 Entrenamiento de Modelo Solución

Este procedimiento consiste en el diseño y construcción del modelo solución como se aprecia en la figura 4.3, es decir el diseño debe contemplar las capas a usarse en el modelo, así como el algoritmo a usarse, el cual es redes neuronales convolucionales descrito en capítulo 2, también se debe definir las métricas para medir su rendimiento como lo son precisión media (AP o MAP) y puntuación F (F Score).

Entrada:

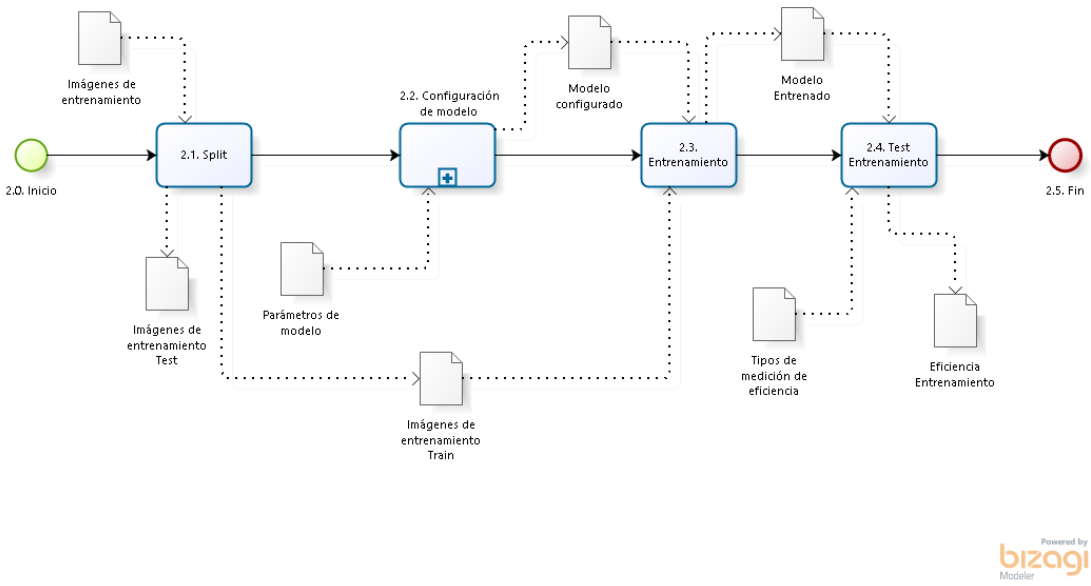


Figura 4.2: Diseño y construcción del modelo solución

4.1.2.1.1 Entrada:

- Imágenes de entrenamiento
- Tipos de medición de eficiencia

4.1.2.1.2 Salida:

- Modelo Entrenado

- Eficiencia Entrenamiento

4.1.2.2 Split

Este procedimiento se realiza la división aleatoria de las imágenes de entrenamiento con un por de 75% para las imágenes del train y el 25% para el test como se aprecia en el cuadro 4.4:

Etiqueta	Imágenes de entrenamiento	Imágenes de train	Imágenes de test
Ave	981	736	245
Gato	783	587	196
Perro	618	464	155
Pez	557	417	139
Flor	563	422	141
Comida	800	600	200
Persona	633	475	158
Reptil	529	397	132
Árbol	45	34	11
Utensilio	671	503	168

4.1.2.2.1 Entrada:

- Imágenes de entrenamiento

4.1.2.2.2 Salida:

- Imágenes de entrenamiento Test

- Imágenes de entrenamiento Train

4.1.2.3 Configuración de modelo

En este procedimiento se realiza la configuración del modelo, para el cual se ha definido 2 capas convolucionales, previas a la red neuronal, la cual contemplará los parámetros cuadro 4.7:

Capa	Tamaño de entrada	Tamaño de característica	Función de activación	Optimizador	Función de pérdida	Métrica
Convolución 1	64, 64, 3	32, 3, 3	RELU (Rectified Linear Unit)	-	-	-
Pooling 1	-	2,2	-	-	-	-
Convolución 2	-	32, 3, 3	RELU (Rectified Linear Unit)	-	-	-
Pooling 2	-	10	-	-	-	-
Capa neuronal de entrada	128	-	-	-	-	-
Capa neuronal de salida		10		-	-	-
Compilar modelo	-	-	-	ADAM	Binary Crossentropy	Accuracy

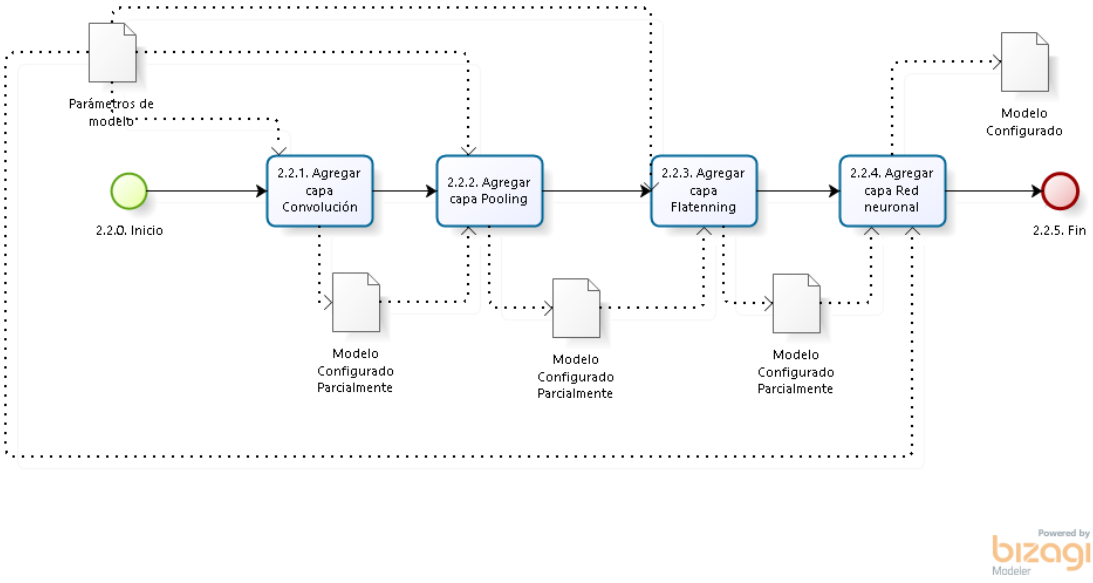


Figura 4.3: Configuración del modelo solución

4.1.2.3.1 Entrada:

- Parámetros de modelo
    - Función de pérdida
    - Función de activación
    - Tamaño del pool
    - Neuronas capa de entrada
    - Neuronas capa de salida (número de clases)

4.1.2.3.2 Proceso:

- Se realiza la configuración para el entrenamiento del modelo y se divide en 4 subprocesos:
    - Agregar capa Convolución
    - Agregar capa Pooling
    - Agregar capa Flatenning

- Agregar capa Red neuronal

4.1.2.3.3 Salida:

- Modelo Configurado

4.1.2.4 Agregar capa Convolución

4.1.2.4.1 Entrada:

- Parámetros de modelo

4.1.2.4.2 Proceso:

- Se configura el modelo para la convolución con su parámetro:
  - Función de activación
  - Tamaño de salida
  - Tamaño de entrada

4.1.2.4.3 Salida:

- Modelo parcialmente configurado

4.1.2.5 Agregar capa Pooling

4.1.2.5.1 Entrada:

- Parámetros de modelo
- Modelo Configurado Parcialmente

4.1.2.5.2 Proceso:

- Se configura el modelo con su parámetros:
  - Función de activación
  - Tamaño de pool

4.1.2.5.3 Salida:

- Modelo parcialmente configurado

4.1.2.6 Agregar capa Flatenning

4.1.2.7 Agregar capa Red neuronal

4.1.2.7.1 Entrada:

- Modelo Configurado Parcialmente
- Parámetros de modelo

4.1.2.7.2 Proceso:

- Se configura el modelo para la red neuronal:
    - Neuronas de entrada
    - Neuronas de salida
    - Función de activación

4.1.2.7.3 Salida:

- Modelo configurado

4.1.2.8 Entrenamiento

4.1.2.8.1 Entrada:

- Imágenes de entrenamiento Train
- Modelo configurado

4.1.2.8.2 Proceso:

- Se realiza el entrenamiento del modelo con las imágenes y el modelo configurado.

4.1.2.8.3 Salida:

- Modelo Entrenado

4.1.2.9 Test Entrenamiento

4.1.2.9.1 Entrada:

- Tipos de medición de eficiencia
    - Matriz de confusión
    - F1 Score
    - Mean average precision
- Modelo Entrenado

4.1.2.9.2 Proceso:

4.1.2.9.3 Salida:

- Eficiencia Entrenamiento
    - Matriz de confusión
    - F1 Score
    - Mean average precision

4.1.2.10 Prueba del modelo

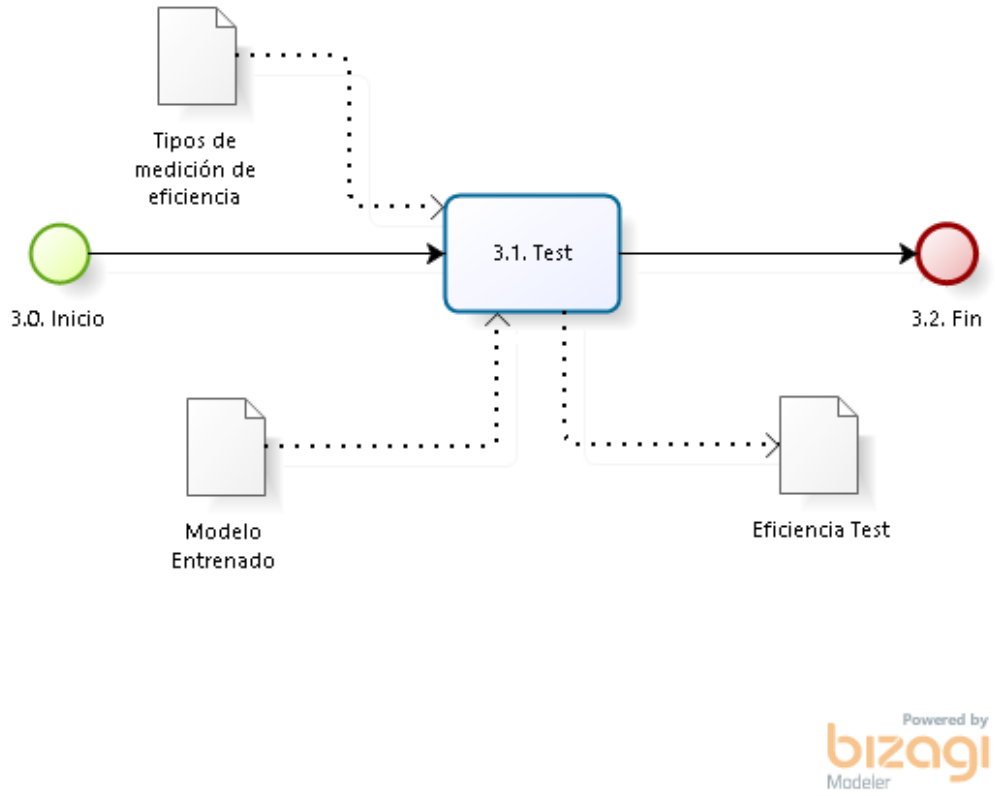


Figura 4.4: Test del modelo solución

4.1.2.10.1 Entrada:

- Tipos de medición de eficiencia
    - Matriz de confusión
    - F1 Score
    - Mean average precision
- Modelo Entrenado

4.1.2.10.2 Proceso:

4.1.2.10.3 Salida:

- Eficiencia Test

- Matriz de confusión
- F1 Score
- Mean average precisión

4.2 RESULTADOS

4.2.1 Pruebas unitarias - Modelo Básico

Las pruebas unitarias realizadas se evaluaron con imágenes descargadas de internet los resultados de algunas de ellos son para el modelo en Jupyter se muestra el top 5 de probabilidades obtenidas para cada imagen

4.2.1.1 Ave

Ave (score=0.55168)
Reptil (score=0.39194)
Utensilio (score=0.10660)
Pez (score=0.06095)
Perro (score=0.05183)

```
jpg
[[5.5168390e-01 3.9665997e-03 5.1828146e-02 6.0953736e-02 4.5701265e-03
  4.0750951e-02 2.0061880e-02 3.9193946e-01 1.4757115e-04 1.0659692e-01]]
[5.5168390e-01 3.9665997e-03 5.1828146e-02 6.0953736e-02 4.5701265e-03
 4.0750951e-02 2.0061880e-02 3.9193946e-01 1.4757115e-04 1.0659692e-01]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utensilio

Evaluation time (1-image): 0.128s

Ave (score=0.55168)
Reptil (score=0.39194)
Utensilio (score=0.10660)
Pez (score=0.06095)
Perro (score=0.05183)
```

4.2.1.2 Comida

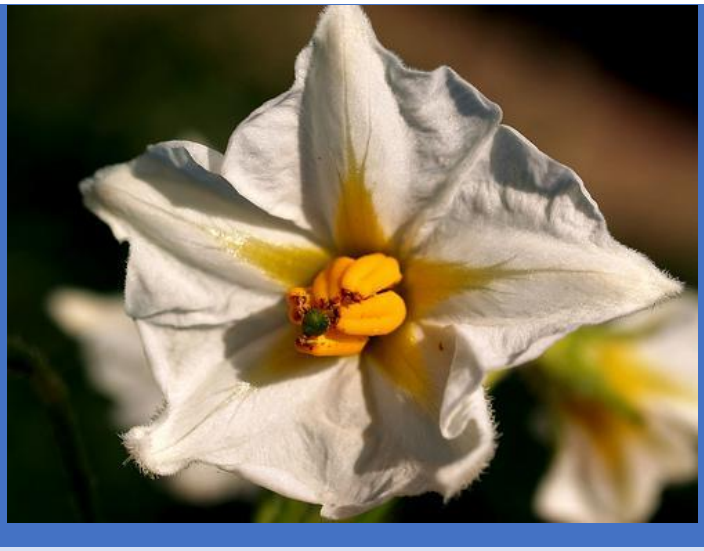
Comida (score=0.99987)
Flor (score=0.00166)
Pez (score=0.00120)
Gato (score=0.00013)
Utensilio (score=0.00013)

```
jpg
[[1.6033649e-05 1.3318658e-04 1.1861324e-05 1.2015700e-03 1.6626120e-03
  9.9987149e-01 2.4139881e-06 7.0899725e-05 2.3275876e-05 1.3056185e-04]]
[1.6033649e-05 1.3318658e-04 1.1861324e-05 1.2015700e-03 1.6626120e-03
 9.9987149e-01 2.4139881e-06 7.0899725e-05 2.3275876e-05 1.3056185e-04]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utensilio']

Evaluation time (1-image): 0.130s

Comida (score=0.99987)
Flor (score=0.00166)
Pez (score=0.00120)
Gato (score=0.00013)
Utensilio (score=0.00013)
```

4.2.1.3 Flor



Flor (score=0.95210)
Comida (score=0.11947)
Pez (score=0.01694)
Persona (score=0.00461)
Gato (score=0.00333)

```
jpg
[[3.2707751e-03 3.3331215e-03 4.6655536e-04 1.6939700e-02 9.5209861e-01
  1.1947250e-01 4.6134293e-03 1.3731122e-03 5.2314831e-06 2.8180884e-04]]
[3.2707751e-03 3.3331215e-03 4.6655536e-04 1.6939700e-02 9.5209861e-01
 1.1947250e-01 4.6134293e-03 1.3731122e-03 5.2314831e-06 2.8180884e-04]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utensilio

Evaluation time (1-image): 0.141s

Flor (score=0.95210)
Comida (score=0.11947)
Pez (score=0.01694)
Persona (score=0.00461)
Gato (score=0.00333)
```

4.2.1.4 Gato

Gato (score=0.98599)
Perro (score=0.05299)
Utensilio (score=0.02251)
Pez (score=0.01433)
Persona (score=0.00624)

```
jpg
[[4.1540265e-03 9.8598516e-01 5.2986711e-02 1.4325857e-02 8.4549189e-05
  3.4093857e-05 6.2426627e-03 1.6541481e-03 2.6858675e-07 2.2513339e-02]]
[4.1540265e-03 9.8598516e-01 5.2986711e-02 1.4325857e-02 8.4549189e-05
 3.4093857e-05 6.2426627e-03 1.6541481e-03 2.6858675e-07 2.2513339e-02]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utensili

Evaluation time (1-image): 0.150s

Gato (score=0.98599)
Perro (score=0.05299)
Utensilio (score=0.02251)
Pez (score=0.01433)
Persona (score=0.00624)
```

4.2.1.5 Perro

Perro (score=0.84811)
Gato (score=0.24164)
Ave (score=0.10057)
Persona (score=0.05738)
Reptil (score=0.01254)

```
jpg
[[1.0056922e-01 2.4163541e-01 8.4810686e-01 9.0175867e-04 4.3958426e-05
  1.3378263e-04 5.7384163e-02 1.2544066e-02 1.2627985e-06 5.0260550e-03]]
[1.0056922e-01 2.4163541e-01 8.4810686e-01 9.0175867e-04 4.3958426e-05
 1.3378263e-04 5.7384163e-02 1.2544066e-02 1.2627985e-06 5.0260550e-03]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utensilio

Evaluation time (1-image): 0.151s

Perro (score=0.84811)
Gato (score=0.24164)
Ave (score=0.10057)
Persona (score=0.05738)
Reptil (score=0.01254)
```

4.2.1.6 Persona

Persona (score=0.36379)
Pez (score=0.11069)
Perro (score=0.09442)
Reptil (score=0.06520)
Ave (score=0.06465)

```
jpg
[[0.06464577 0.01760766 0.09442344 0.11069471 0.00079754 0.02558404
  0.36379236 0.06519583 0.00442761 0.02118007]]
[0.06464577 0.01760766 0.09442344 0.11069471 0.00079754 0.02558404
 0.36379236 0.06519583 0.00442761 0.02118007]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utens

Evaluation time (1-image): 0.124s

Persona (score=0.36379)
Pez (score=0.11069)
Perro (score=0.09442)
Reptil (score=0.06520)
Ave (score=0.06465)
```

4.2.1.7 Pez

Pez (score=0.98997)
Ave (score=0.00803)
Utensilio (score=0.00396)
Persona (score=0.00035)
Comida (score=0.00017)

```
jpg
[[8.0265999e-03 1.1217594e-04 1.9073486e-06 9.8996550e-01 2.8342009e-05
  1.7249584e-04 3.5285950e-04 6.8545341e-06 3.3013993e-07 3.9638001e-03]]
[8.0265999e-03 1.1217594e-04 1.9073486e-06 9.8996550e-01 2.8342009e-05
 1.7249584e-04 3.5285950e-04 6.8545341e-06 3.3013993e-07 3.9638001e-03]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utens

Evaluation time (1-image): 0.128s

Pez (score=0.98997)
Ave (score=0.00803)
Utensilio (score=0.00396)
Persona (score=0.00035)
Comida (score=0.00017)
```

4.2.1.8 Reptil

Reptil (score=0.61837)
Flor (score=0.06858)
Pez (score=0.05454)
Ave (score=0.01784)
Gato (score=0.01254)

```
jpg
[[1.7840058e-02 1.2536079e-02 1.2061894e-03 5.4537535e-02 6.8581969e-02
  3.5283566e-03 2.6085079e-03 6.1837399e-01 3.1995292e-05 3.0951873e-07]]
[1.7840058e-02 1.2536079e-02 1.2061894e-03 5.4537535e-02 6.8581969e-02
 3.5283566e-03 2.6085079e-03 6.1837399e-01 3.1995292e-05 3.0951873e-07]
['Ave', 'Gato', 'Perro', 'Pez', 'Flor', 'Comida', 'Persona', 'Reptil', 'Arbol', 'Utensilio

Evaluation time (1-image): 0.151s

Reptil (score=0.61837)
Flor (score=0.06858)
Pez (score=0.05454)
Ave (score=0.01784)
Gato (score=0.01254)
```

4.2.2 Pruebas unitarias – Modelo Multicapa

4.2.2.1 Jabón Palmolive Sensación Humectante

Modelo Nivel 1	Modelo Nivel 2	Modelo Nivel 3
Jabon (score=0.99999)	Palmolive (score=0.99189)	Sensacion Humectante (score=0.99904)
Shampoo (score=0.00567)	Lux (score=0.00000)	Suavidad Relajante (score=0.01113)
Pasta Dental (score=0.00000)	Protex (score=0.00000)	Delicada Exfoliacion (score=0.00012)
Gaseosa (score=0.00000)	Otros (score=0.00000)	Otros (score=0.00000)
Cafe (score=0.00000)	Camay (score=0.00000)	
Jabon Palmolive Sensacion Humectante		

```
Evaluation time (1-image): 0.155s

Jabon (score=0.99999)
Shampoo (score=0.00567)
Pasta Dental (score=0.00000)
Gaseosa (score=0.00000)
Cafe (score=0.00000)
Jabon

Evaluation time (1-image): 0.120s

Palmolive (score=0.99189)
Lux (score=0.00000)
Protex (score=0.00000)
Otros (score=0.00000)
Camay (score=0.00000)

Evaluation time (1-image): 0.121s

Sensacion Humectante (score=0.99904)
Suavidad Relajante (score=0.01113)
Delicada Exfoliacion (score=0.00012)
Otros (score=0.00000)

El resultado es: Jabon Palmolive Sensacion Humectante
```

4.2.2.2 Gaseosa Coca Cola

Modelo Nivel 1	Modelo Nivel 2	Modelo Nivel 3
Gaseosa (score=1.00000)	**Coca Cola (score=0.99896)**	
Cafe (score=0.00000)	**KR (score=0.00036)**	

Pasta Dental (score=0.00000)	**Otros (score=0.00001)**	
Otros (score=0.00000)	**Sprite (score=0.00000)**	
Agua Mineral (score=0.00000)	**Pepsi (score=0.00000)**	
Gaseosa Coca Cola		

```
Evaluation time (1-image): 0.157s

Gaseosa (score=1.00000)
Cafe (score=0.00000)
Pasta Dental (score=0.00000)
Otros (score=0.00000)
Agua Mineral (score=0.00000)
Gaseosa

Evaluation time (1-image): 0.128s

Coca Cola (score=0.99896)
KR (score=0.00036)
Otros (score=0.00001)
Sprite (score=0.00000)
Pepsi (score=0.00000)

El resultado es: Gaseosa Coca Cola
```

4.2.2.3 Shampoo Pantene Hidratación Extrema

Modelo Nivel 1	Modelo Nivel 2	Modelo Nivel 3
Shampoo (score=0.99975)	**Pantene (score=1.00000)**	**Hidratacion Extrema (score=1.00000)**
Pasta Dental (score=0.00196)	**Dove (score=0.00000)**	**Otro (score=0.00000)**
Cafe (score=0.00003)	**Savital (score=0.00000)**	**Restauracion (score=0.00000)**
Agua Mineral (score=0.00002)	**Sedal (score=0.00000)**	**Fuerza y Reconstruccion (score=0.00000)**
Jabon (score=0.00001)	**Pert Plus (score=0.00000)**	

Shampoo Pantene Hidratacion Extrema

```
Evaluation time (1-image): 0.154s

Shampoo (score=0.99975)
Pasta Dental (score=0.00196)
Cafe (score=0.00003)
Agua Mineral (score=0.00002)
Jabon (score=0.00001)
Shampoo

Evaluation time (1-image): 0.123s

Pantene (score=1.00000)
Dove (score=0.00000)
Savital (score=0.00000)
Sedal (score=0.00000)
Pert Plus (score=0.00000)

Evaluation time (1-image): 0.146s

Hidratacion Extrema (score=1.00000)
Otro (score=0.00000)
Restauracion (score=0.00000)
Fuerza y Reconstruccion (score=0.00000)

El resultado es: Shampoo Pantene Hidratacion Extrema
```

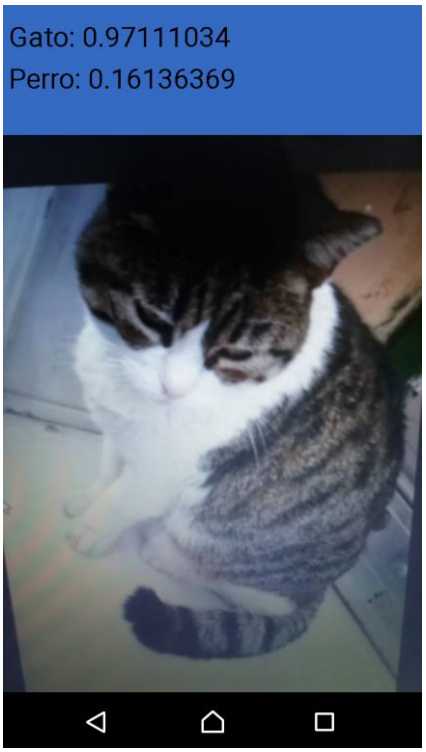
4.2.3 Resultados Aplicativo

En el aplicativo de lo capturado en tiempo real se da una predicción mediante voz y se muestra los resultados de mayo probabilidad.

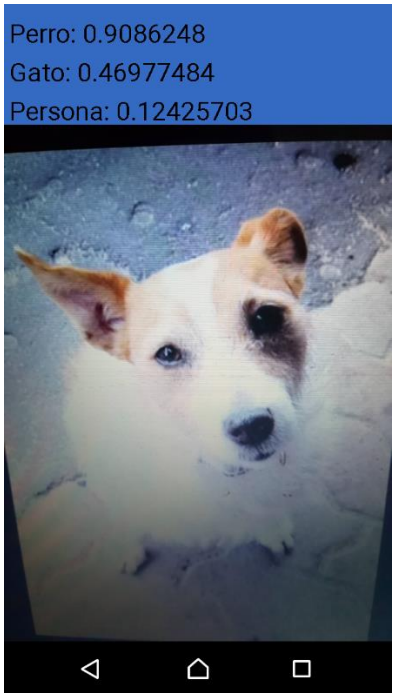
4.2.3.1 Persona

4.2.3.2 Comida

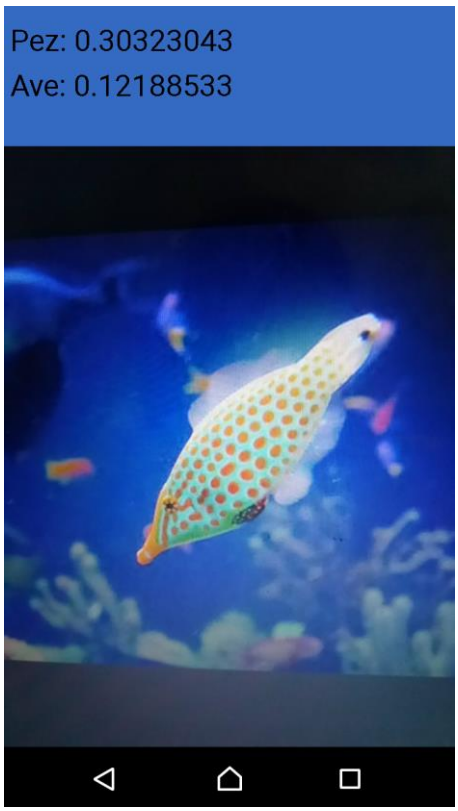
4.2.3.3 Gato



4.2.3.4 Perro



4.2.3.5 Pez



4.3 ANÁLISIS DE RESULTADOS

Se obtuvo como resultados al probar 320 imágenes de cada clase:

4.3.1 Matriz confusión:

		Valor predicho									
		Ave	Gato	Perro	Pez	Flor	Comida	Persona	Reptil	Árbol	Utensilio
Valor Real	Ave	281	4	5	4	4	3	5	7	0	7
	Gato	24	227	7	8	6	9	14	8	0	17
	Perro	28	6	226	8	8	3	14	14	0	13
	Pez	28	2	2	244	3	9	6	9	0	17
	Flor	6	0	0	0	308	4	0	1	0	1
	Comida	0	0	0	4	7	308	0	0	0	1
	Persona	9	1	3	6	2	4	273	6	1	15
	Reptil	17	2	5	9	12	7	4	253	0	11
	Árbol	10	0	0	10	35	15	0	0	230	0
	Utensilio	10	1	4	3	1	5	8	5	0	283

4.3.2 Medidas de eficiencia:

	Precision	Recall	F1 Score
Ave	0.680387	0.878125	0.766712
Gato	0.934156	0.709375	0.806394
Perro	0.896825	0.70625	0.79021
Pez	0.824324	0.7625	0.792208
Flor	0.797927	0.9625	0.872521
Comida	0.839237	0.9625	0.896652
Persona	0.842593	0.853125	0.847826
Reptil	0.834984	0.790625	0.812199
Árbol	0.995671	0.766667	0.86629

Utensilio	0.775342	0.884375	0.826277
Average F1 Score	-	-	82.7729
Mean Average Precision	84.21448	-	-

CAPÍTULO V CONCLUSIONES

- El modelo multicapa demuestra mejor eficiencia para el método de redes neuronales convolucionales, con una base de datos no tan grande.
- El etiquetado de objetos automático es una herramienta que permite a las personas con discapacidad visual a tener una mejor calidad de vida y poder realizar actividades de mejor manera.
- El uso de redes neuronales convolucionales o CNN (Convolutional Neural Network) es un método efectivo para poder etiquetar imágenes de objetos.
- Para el uso de redes neuronales convolucionales se debe tener gran cantidad de imágenes de entrenamiento, ya que es un método que se basa en un ajuste de este.

CAPÍTULO VI RECOMENDACIONES

- El modelo está basado en 10 etiquetas, sería bueno que se aumente esta cantidad, para poder realizar otros usos, como herramientas para compras u otros.
- La recolección de imágenes para entrenamiento es un paso de vital importancia, por ende, es recomendable que por etiqueta se tenga un aproximado de 1000 imágenes para tener un buen rendimiento.
- El modelo multicapa mejora la eficacia de predicción, sin embargo, al agregarse más capas, se debe evaluar el número de capas la imagen, además se debe entrenar una mayor cantidad de modelos, el tiempo de predicción es mayor al que si solo fuera un modelo.

CAPÍTULO VII REFERENCIAS BIBLIOGRÁFICAS

[1] X.; Ma Y.; Wright-J. CandA˜ ¨s, E.J.; Li. Robust principal component analysis page e11, 2011.

[2] C. Fellbaum. Wordnet: An electronic lexical database. Bradford Books, 1998.

[3] Jianlong Fu and Yong Rui. Advances in deep learning approaches for image tagging. APSIPA Transactions on Signal and Information Processing, 6:e11, 2017.

[4] Y.;Mei T.; Wang J.; Lu-H.; Rui Y. Fu, J.;Wu. Relaxing fromvocabulary: robust weakly- supervised deep learning for vocabulary-free image tagging. IEEE Int. Conf. on Computer Vision, 2015.

[5] Richard Socher Li-Jia Li Kai Li Jia Deng, Wei Dong and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. IEEE Transactions on Multimedia, pages 248–255, 2009.

[6] I.; Hinton G.E Krizhevsky, A.; Sutskever. Imagenet classifiGatoion with deep convolutional neural networks, in advances in neural information processing systems. IEEE Transactions on Multimedia, 2012.

[7] X.; Tang X. Luo, P.;Wang. Hierarchical face parsing via deep learning. IEEE Transactions on Multimedia, pages 2480–2487, 2012.

[8] R. Sukhbaatar, S.; Fergus. Learning from noisy labels with deep neural networks. IEEE Transactions on Multimedia, 2014.

[9] Franco Javier Ascarza Mendoza. Segmentación automática de textos, mediante redes neuronales convolucionales en imágenes documentos históricos. PUCP Escuela de posgrado, 2018.